
화학 공정 시스템 내 위험 규명을 위한 텍스트마이닝



7조 강영은 류건희 이건학 이지선

TABLE OF CONTENTS

1	프로젝트 배경
2	데이터 수집 및 전처리, 탐색
3	전체 데이터 분석
4	군집별 데이터 분석
5	프로젝트 결론

FLOW CHART



현재의 방식 & 데이터 마이닝 문제

현재의 화학 공정 시스템 내 위험 규명을 위해서는 전문가들의 토론, 과거 사례를 참고하는 정도의 방식이 이용되고 있음

그러나, 이러한 방식은 여러 문제를 가지고 있음

1. 많은 시간 / 비용 소모
2. 예상치 못한 위험요소 존재 가능



- 과거 사례에 대한 데이터 마이닝을 통한 잠재적 위험 분석
- Substance와 Process Unit, Final Event의 관계를 알아보고자 함

데이터 수집 및 전처리, 탐색 사용한 데이터

- RISCAD (Relational System Information for Chemical Accidents Database)의 5113개 사고 data
- 보고서 형식의 Text Data
- 총 17개의 항목 중 결측치가 많고 중요하지 않은 항목은 제외 후 Damage, Involved Substance, Equipment, Final event, Cause 등의 주요 항목 10개를 선정하여 분석

Relational

Information

System

for

RISCAD

Accidents

Database

Chemical

Accidents

Database

Home

Case Search

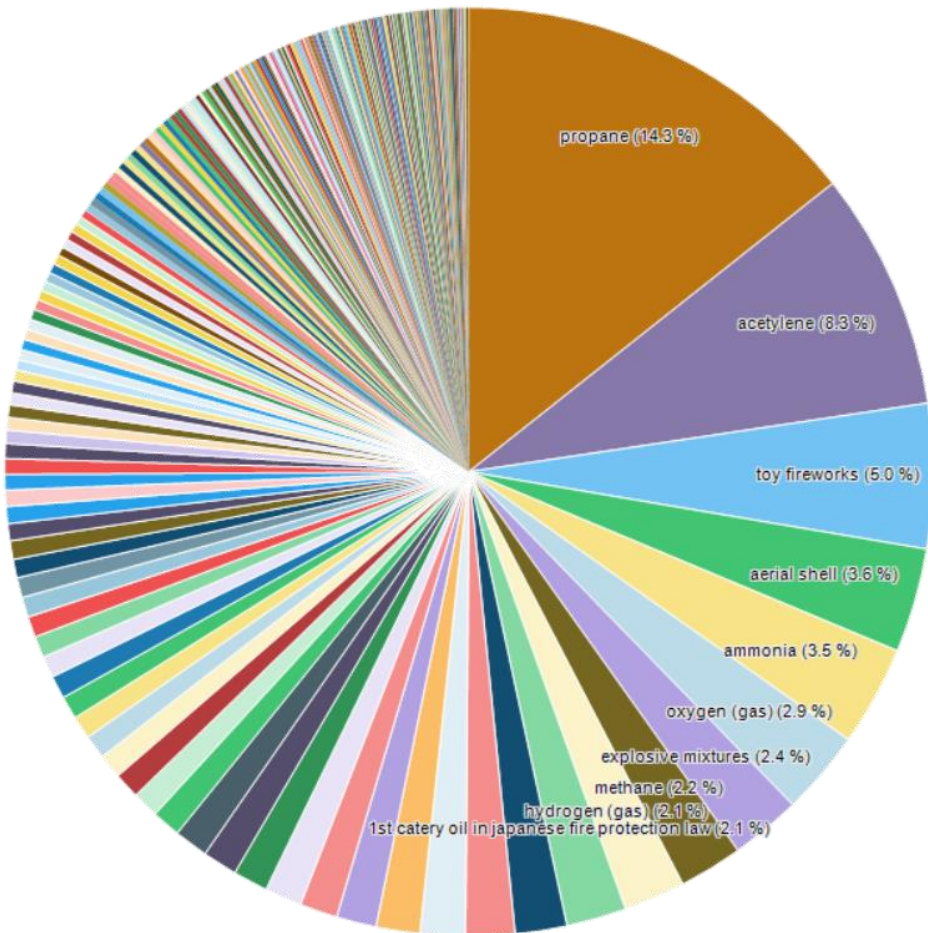
Material Search

How To Use

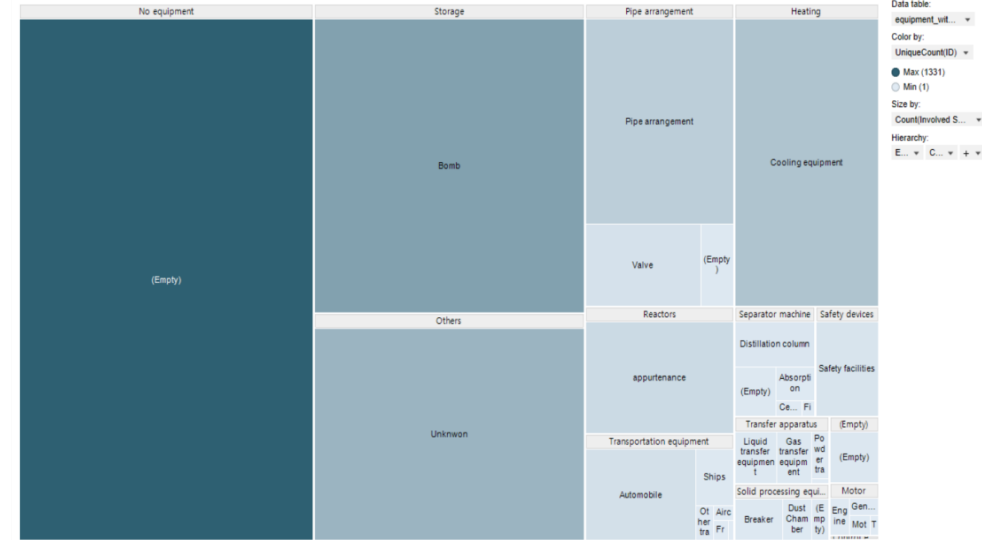
Characteristics

ID	1	Date	1988/06/03	
Name	Explosion in a hazardous material manufacturer (a vacuum distillation column for high boiling point compounds)		Time	
Country	Place name		Fukushima / Iwaki City	
Industry	Manufacture of chemical and allied products			
Final Event	Fire Explosion / Explosive decomposition			
Human Damage	Fatality	Injury	Toxic	Discomfort
Damage*	Physical Damage / Building / Accident institution Physical Damage / Apparatus / Accident equipment			
Cost	48.4 million Yen			
Summary	A fire occurred at a vacuum distillation column for refining crude phenol. At the final stage of the process, loading of crude phenol was stopped and 800 liters of phenol in the evaporator was being taken out. When it reached about 500 liters of level, the process temperature rose from 160°C to 230°C. High boiling point substances that had been mixed with raw material in the evaporator started exothermic decomposition and discharged a large quantity of gas, causing the pressure in the evaporator to go up rapidly, and led to the rupture of the evaporator. This produced sparks, which ignited flammable leaking gases.			

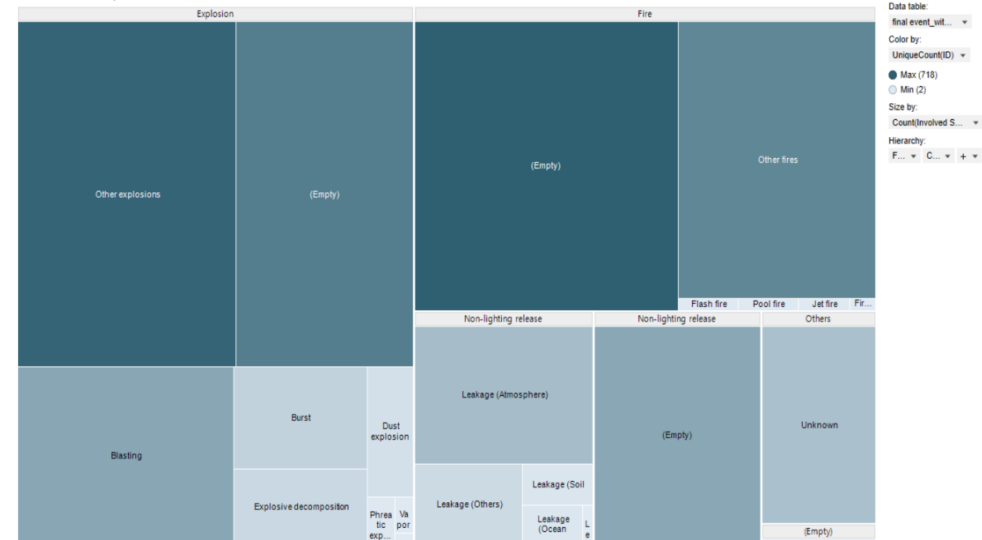
데이터 수집 및 전처리, 탐색 주요 항목 데이터 탐색



Involved Substance per Equipment, Column 25



Involved Substance per Final Event, Column 25



주요지표 연관 분석

- Substance, Process Unit, Final Event 세 항목의 Noun과 Verb에 대해서 연관 분석 시행
- Minimum Support: 0.01, Confidence: 0.3, Lift: 5에 대해 Rule 765개 도출
- 하지만 무의미한 결과가 많이 섞여 있어 의미 있는 rule을 찾기 어려움
 - Bigram의 두 단어가 연관성이 있다고 나옴
Ex) 화학물질 aerial shell : {'aerial'} → {'shell'}, confidence=1.0, lift=36.3
 - Process Unit, Final Event의 대분류와 소분류의 단어가 연관성이 있다고 나옴
Ex) Process Unit인 Storage/Bomb/Gas holder: {'holder'} → {'bomb'}, confidence=1.0, lift=6.530012771392081)

주요지표 연관 분석

{ 'ammonium' } → { 'nitrate' }

support=0.017993350283590847, confidence=0.9583333333334, lift=44.5450757576

{ 'explosive' } → { 'nitrate' } / { 'nitrate' } → { 'explosive' }

support= 0.017798, confidence=0.7777777778, lift=36.15252526 / confidence=0.8272727273

{ 'explosive', 'ammonium' } → { 'nitrate' }

support= 0.01779777038920399, confidence=1.0, lift=46.481818181818184

⇒ **질산암모늄** - 용도 1) 비료. 2) **폭발**. 3) 한제. 4) 효모 배양의 양분. 5) 산화이질소의 제조

{ 'gas', 'accident' } → { 'acetylene' }

support=0.015450811656561706, confidence=0.33054393305439334, lift=5.248668104680475

{ 'apparatus', 'bomb' } → { 'acetylene' }

support= 0.018188930177977704, confidence=0.33333333333333337, lift=5.29296066252588

{ 'gas', 'damage' } → { 'acetylene' }

support= 0.020731468805006845, confidence=0.32415902140672787, lift=5.147282846126085

⇒ **아세틸렌** : 토치로 산소와의 혼합기를 만들어 연소하면 **최고 3500°C의 고온**이 되기 때문에 용접에 있어서 열원으로 사용. **아세틸렌과 공기를 혼합시킨 기체는 폭발의 위험이 있음.**

주요지표 연관 분석

{‘propane’, ‘bomb’} → {‘gas’} / {‘gas’, ‘propane’} → {‘bomb’}

support=0.031292783, confidence=0.96385542, lift=5.85296054 / confidence=0.97560976, lift=6.37074417

{‘propane’, ‘bomb’} → {‘storage’} / {‘propane’, ‘storage’} → {‘bomb’}

support= 0.0324662624, confidence=1.0, lift=6.530012771 / confidence=1.0, lift=6.530012771

⇒ **프로판** : 불이 매우 잘 붙고, 일단 불이 붙으면 상당한 속도로 연소. 이러한 **폭발적인 연소**를 이용하여 자동차 엔진에 쓰임. 일단 산소와의 결합이 시작되면 그 발열로 인해 연쇄적으로 주위의 연료 분자가 산소와 반응. **가스 폭발**은 프로판 가스가 새어 나와 산소와 뒤섞인 상태에서 발생.

{‘oxygen’, ‘equipment’} → {‘gas’}

support=0.010756894191277136, confidence=1.0, lift=6.072446555819478

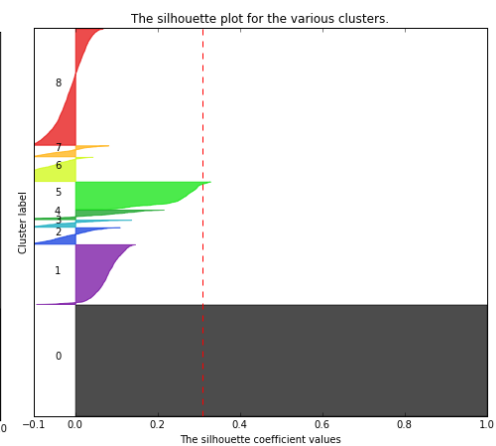
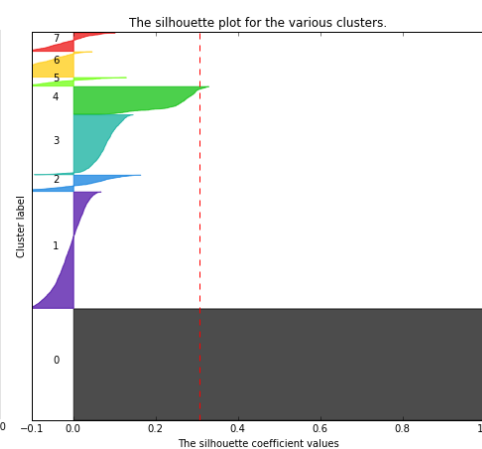
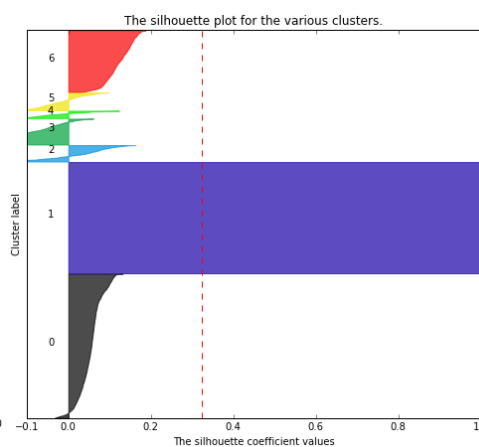
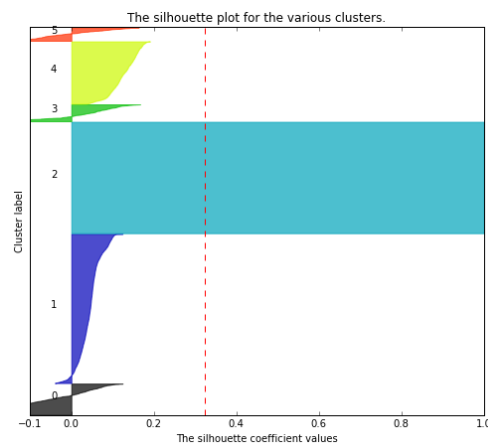
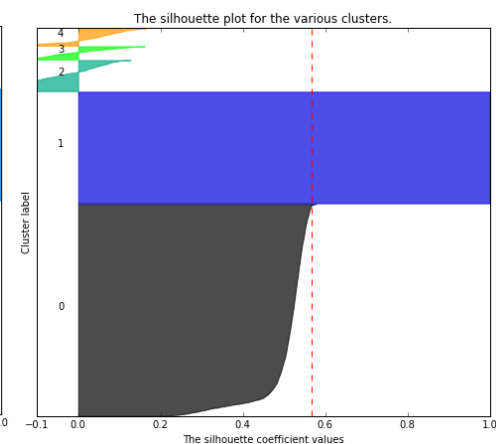
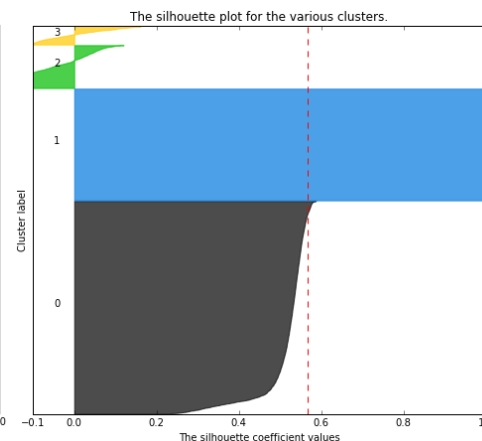
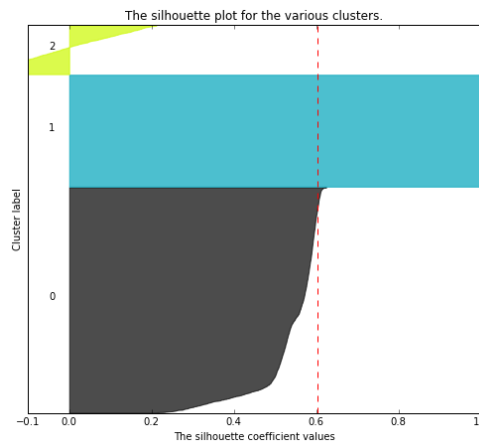
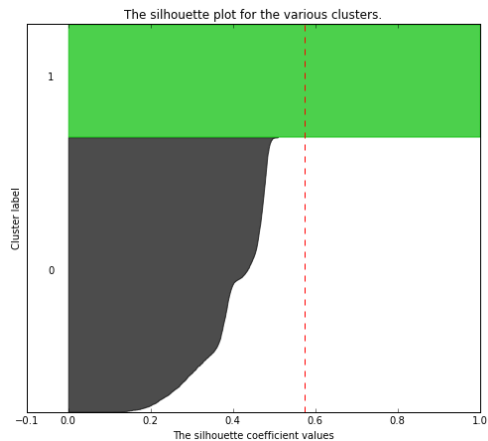
⇒ **산소** : 산소 기체 자체는 타지 않지만 다른 물질이 타는 것을 도우며(**조연성**), **반응성이 매우 커서** 할로젠을 제외한 거의 모든 원소와 반응하여 산화물을 만든다.

가장 많이 사고를 일으키는 폭발성 물질 : **nitrate ammonium, acetylene, propane + oxygen**(폭발성X, 조연성O)

(전체 문서에 포함된 폭발성 물질 : 214 종류)

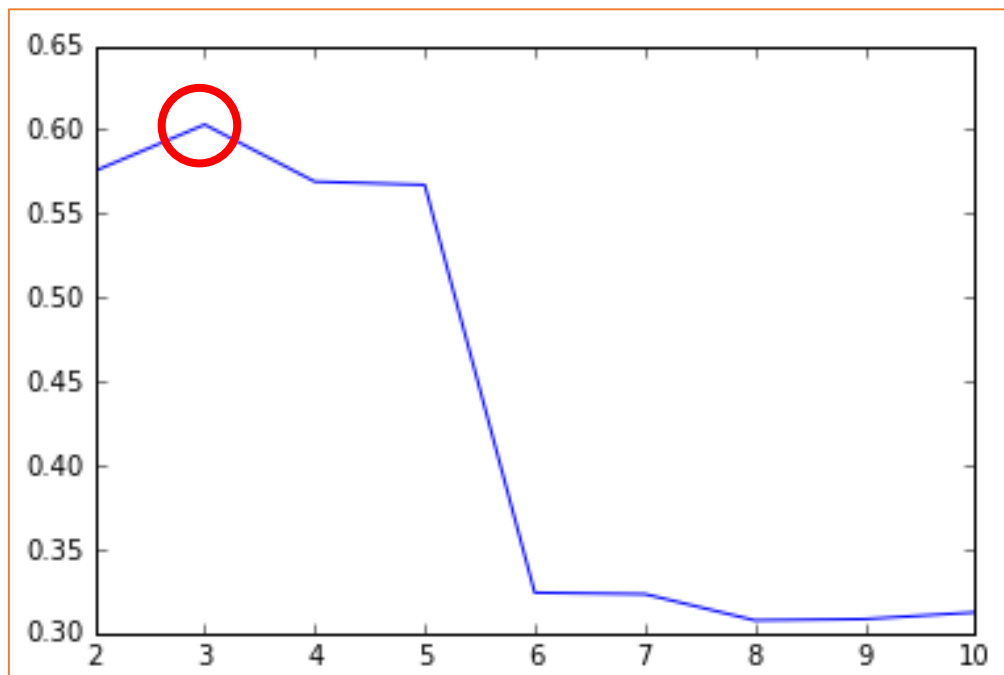
→ 한계 : 물질과 Process Unit 간의 상호작용으로 인한 사고 발생에 대한 association rule은 발견하지 못함

군집별 데이터 분석 클러스터링



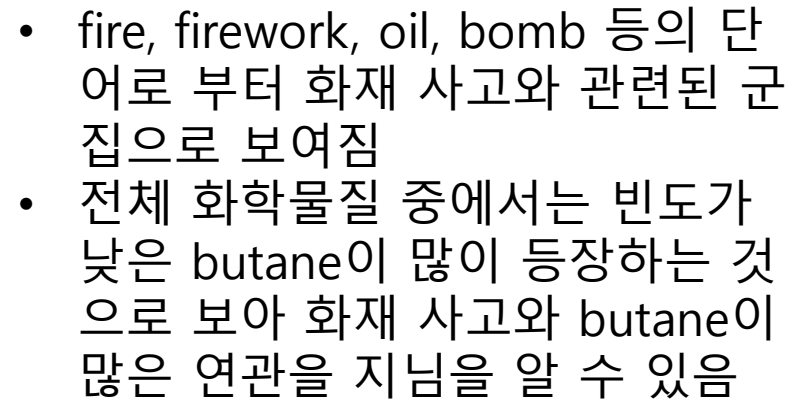
군집별 데이터 분석 클러스터링

```
For n_clusters = 2 The average silhouette_score is : 0.575339281175
For n_clusters = 3 The average silhouette_score is : 0.603035603432
For n_clusters = 4 The average silhouette_score is : 0.567916010464
For n_clusters = 5 The average silhouette_score is : 0.565806985755
For n_clusters = 6 The average silhouette_score is : 0.323829144028
For n_clusters = 7 The average silhouette_score is : 0.323260124504
For n_clusters = 8 The average silhouette_score is : 0.307886325377
For n_clusters = 9 The average silhouette_score is : 0.308226266657
For n_clusters = 10 The average silhouette_score is : 0.31152816032
```



Silhouette 값이 가장 큰
3으로 클러스터 수 결정

Word Cloud - TF



- 화재성이 있는 firework, dioxane 등의 단어가 부각됨
 - dioxane의 경우 화재성과 폭발성을 둘 다 지니고 있음에도 화재성 군집에만 특히 나타남
- 단순히 **물질의 성질만 가지고 관련 사고를 추측**해서는 안 됨

클러스터별 연관 분석 – cluster 0

{'chloromethane'} → {'appurtenance'}

support=0.018234966592427616, confidence=1.0, lift=19.900277008310248

{'chloromethane'} → {'reactor'} / {'reactor'} → {'chloromethane'}

support=0.01823496659243, confidence=1.0, lift=19.90027700831 / confidence=0.362880886426

{'lighting', 'appurtenance'} → {'chloromethane'}

support=0.018234966592427616, confidence=0.7660818713450291, lift=42.01169590643275

⇒ **염화메틸** - 용도 1) 냉동기용 냉매 2) 저온용 용제, 추출제 3) 유기 합성의 메틸화제 및 중간체

reactor - 석유화학공업이나 메탄올, 암모니아의 합성화학공업 등의 촉매화학 반응용에 사용

{'chloroethylene'} → {'arrangement'}

support= 0.012806236080178173, confidence=0.9583333333333333, lift=9.303603603603603

{'chloroethylene'} → {'fire'}

support= 0.013084632516703786, confidence=0.9791666666666666, lift=6.757284662183797

{'chloroethylene', 'arrangement'} → {'arrangement'}

support= 0.012667037861915368, confidence=0.9891304347826088, lift=6.826045190661153

⇒ **염화비닐** - 용도 폴리염화비닐, 염화비닐-아세트산비닐 혼성 중합체, 염화비닐리덴, 염화비닐 혼성 중합체의 합성에 사용

클러스터별 연관 분석 – cluster 0

{‘burst’} → {‘tetrahydrofuran’}

support= 0.017797770, confidence=0.67676767, lift=24.18855219 / confidence=1.0

{‘burst’, ‘equipment’} → {‘tetrahydrofuran’}

support= 0.027978841870824053, confidence=0.9136363636363636, lift=32.654545454545456

{‘burst’, ‘explosion’} → {‘tetrahydrofuran’}

support= 0.027978841870824053, confidence=0.6767676767676767, lift=24.188552188552187

⇒ **테트라히드로푸란** - 용도 1) 천연 및 합성 수지의 양호한 용제 2) 유기 약품의 제조 원료

{‘oxygen’} → {‘appurtenance’}

support=0.012249443207126948, confidence=0.8712871287128713, lift=17.338855215161406

{‘oxygen’} → {‘fire’}

support=0.012806236080178173, confidence=0.9108910891089109, lift=6.286111031852465

{‘oxygen’} → {‘gas’}

support=0.01405902004454343, confidence=1.0, lift=5.8692810457516345

⇒ **산소** : 산소 기체 자체는 타지 않지만 다른 물질이 타는 것을 도우며(**조연성**), **반응성이 매우 커서** 할로젠을 제외한 거의 모든 원소와 반응하여 산화물을 만든다.

클러스터별 연관 분석 – Cluster 0

{'chlorine'} → {'gas'}

support= 0.013780623608017817, confidence=0.8761061946902654, lift=5.142113482561165

⇒ **염소** : 염소는 광합성에서 산소 발생을 수반하는 광화학반응에 망간과 함께 **촉매**적으로 작용

{'propane'} → {'fire'}

support= 0.03507795100222717, confidence=0.8689655172413794, lift=5.99678028420948

⇒ **프로판** : 불이 매우 잘 붙고, 일단 불이 붙으면 상당한 속도로 연소. 이러한 **폭발적인 연소**를 이용하여 자동차 엔진에 쓰임. 일단 산소와의 결합이 시작되면 그 발열로 인해 연쇄적으로 주위의 연료 분자가 산소와 반응. **가스 폭발**은 프로판 가스가 새어 나와 산소와 뒤섞인 상태에서 발생.

{'column', 'explosion'} → {'distillation'}

support= 0.05317371937639198, confidence=1.0, lift=18.56330749354005

{'column', 'machine'} → {'distillation'}

support= 0.053869710467706014, confidence=1.0, lift=18.56330749354005

⇒ **증류** : 어떤 용질이 녹아 있는 용액을 가열하여 얻고자 하는 액체의 끓는점에 도달하면 기체상태의 물질이 생긴다. 이를 다시 냉각시켜 액체상태로 만들고 이를 모아 순수한 액체를 얻는 과정

클러스터별 연관 분석 – Cluster 2

{ 'gas', 'bomb', 'propane' } → { 'holder' },

support= 0.021993318485523387, confidence=1.0, lift=13.554716981132076

{ 'non', 'lighting', 'bomb' } → { 'propane' },

support=0.012388641425389755, confidence=0.489010989010989, lift=6.729990316197212)

{ 'cooling', 'equipment', 'explosion' } → { 'propane' } / { 'propane', 'equipment', 'explosion' } → { 'cooling' }

support=0.011971046770601337, confidence=0.44102564102564107/0.826923076923077,
lift=6.069594262697712 /15.470352564102566

⇒ **프로판** – 불이 매우 잘 붙고, 일단 불이 붙으면 상당한 속도로 연소.

- Process Unit 중 특히 cooling equipment가 사고와 연관이 높음
- Non-lighting release 사고와 연관이 높음

{ 'fire', 'acetylene' } → { 'bomb' },

support= 0.0263084632516703, confidence=0.8709677419354839, lift=9.283430649947354

{ 'fire', 'bomb' } → { 'acetylene' },

support=0.0263084632516703, confidence=0.5981012658227848, lift=16.589804994868285

⇒ **아세틸렌** – 아세틸렌과 공기를 혼합시킨 기체는 폭발의 위험이 있음.

- Process Unit 중 bomb에서 사고가 많이 발생

클러스터별 연관 분석 – Cluster 2

{'nitrate'} → {'blasting'},

support=0.01183184855233853, confidence=0.8947368421052632, lift=8.805191059841384

{'blasting', 'ammonium'} → {'nitrate'}

support=0.01183184855233853, confidence=0.9883720930232557, lift=74.74173806609546

⇒ **질산암모늄** - 용도 1) 비료. 2) **폭발**. 3) 한제. 4) 효모 배양의 양분. 5) 산화이질소의 제조
- explosion 중 특히 blasting 사고와 연관이 높음

{'ammonia'}) → {'lighting'},

Support=0.016564587973273943, confidence=0.7300613496932515, lift=7.9950620978602425)

{'lighting', 'ammonia'} → {'release'} / {'release', 'ammonia'} → {'lighting'}

support= 0.016564587973273943, confidence=1.0, lift=10.951219512195122

⇒ **암모니아** - 석유 등의 연료와 함께 있을 때에는 강력한 산화제로 작용하여 **폭발**
- Non-lighting release 사고와 연관이 높음

프로젝트 결론

- 본 프로젝트에서는 화학 공정 내 사고의 특징에 대해서 분석하고 가장 중요한 항목인 Substance, Process Unit, Final Event의 관계에 대해서 알아보고자 함
- 우선 전체 문서에 대해서 Wordcloud를 만들고 연관 분석을 시행한 결과, 가장 많이 사고를 일으키는 물질들은 발견하였으나 목표했던 rule을 찾아내는 것에는 **어려움이 존재**
- 따라서, 군집 분석을 통해 사고를 분류한 뒤 Wordcloud를 만들고 연관 분석을 시행
- Wordcloud의 결과, 화재성, 폭발성과 같은 물질의 성질이 동일하게 화재, 폭발과 같은 사고로 **이어지지 않음**을 알 수 있었음
- 연관 분석 결과, 각 군집의 특성에 맞는 rule이 나왔으며 각 물질과 **구체적인** 사고, 각 물질과 **구체적인** Process Unit과의 관계를 알 수 있었음
- 본 프로젝트는 화학 공정의 **사고**를 다루기 때문에 데이터가 많지 않았지만 더 많은 데이터를 얻는다면 더 신뢰성 있는 rule을 얻을 수 있을 것으로 기대 됨

**이상으로 발표 마치겠습니다.
감사합니다.**
